

ABSTRACT

A method and apparatus are disclosed for alleviating congestion and overload in a distributed call-processing system interconnected through a packet based network, such as an IP or an ATM network. The illustrative IP network includes a plurality of end terminals (ETs) and distributed call processors (CPs). When an end terminal (ET) wants to place a call, the end terminal (ET) send a call set up message to a call processor (CP). According to an aspect of the invention, the call processor will determine whether to process the request or to forward the request to another call processor. Generally, the call processor will declare an overload condition if sufficient resources (such as processing or memory resources) are not available to process a given call. If a call processor determines that it is too congested to process a call, the call processor enters an overload condition, selects an alternate call processor and forwards the request to the alternate call processor. A given call processor (CP) implicitly announces its overload condition to another call processor by virtue of the forwarded congestion message. Each call processor (CP) maintains an ordered list of call processors that indicates whether or not each call processor (CP) is overloaded. The present invention attempts to distribute forwarded congestion messages among all of the available alternate call processors (CPs), for example, using a last message sent (LMS) flag. Generally, a call processor (CP) in an overload condition will not forward another congestion message to a call processor (CP) having its last message sent (LMS) flag set unless there are no other call processors (CPs) available. The congested call processor attaches a call processor identifier to the forwarded congestion message, indicating to the recipient call processor that the congested call processor is in an overload condition. Thus, a forwarded congestion message will cause the recipient call processor to set a flag, for example, in the ordered list of call processors (CPs), indicating that the congested call processor is congested. In one embodiment, each congestion flag has an associated timer that causes the flag to expire (or reset) after a predefined time interval that permits the congested call processor to recover from the overload condition.